

E. COMUNICACIÓN CIENTÍFICA

La Declaración de San Francisco (DORA) y la mala bibliometría

The San Francisco Declaration (DORA) and bad bibliometrics

Isidro F. Aguillo

Aguillo, Isidro F. (2015). "La Declaración de San Francisco (DORA) y la mala bibliometría". *Anuario ThinkEPI*, v. 9, pp. 183-188.

<http://dx.doi.org/10.3145/thinkepi.2015.43>

Publicado en *IweTel* el 19 de febrero de 2015



Resumen: La Declaración de San Francisco (DORA) no sólo critica el uso abusivo del factor de impacto de Garfield como herramienta para la evaluación de la actividad científica, sino que cuestiona de forma implícita el análisis de citas en particular y la práctica bibliométrica en general. Se reivindica el trabajo de la comunidad bibliométrica profesional, identificando errores técnicos y prácticas comerciales por parte de las empresas proveedoras de datos bibliométricos responsables en buena medida de los problemas estructurales, mientras que alerta de una situación coyuntural fruto de la generalización de una bibliometría fácil que habría que corregir con una declaración de buenas prácticas ahora en elaboración por un grupo de bibliómetras.

Palabras clave: Declaración de San Francisco (DORA); Bibliometría; Evaluación de la actividad científica; Factor de impacto; Uso abusivo de indicadores; Citas bibliográficas.

Abstract: The San Francisco Declaration on Research Assessment (DORA) not only criticizes the misuse of Garfield's impact factor as a tool for the evaluation of scientific activity, but implicitly questions citation analysis in particular and bibliometric practice in general. The work of the professional bibliometric community is defended, identifying technical errors and commercial practices of the companies providing bibliometric data, which are largely responsible for the structural problems, while warning of a conjunctural situation that results from the generalization of "easy bibliometrics". This situation should be corrected with a future statement of good practice now under development by a group of bibliometrists.

Keywords: Declaration of San Francisco on Research Assessment (DORA); Bibliometrics; Evaluation of scientific activity; Impact factor; Abusive use of indicators; Bibliographic citations.

Colectivos contrarios a las métricas

La publicación de la llamada *Declaración de San Francisco*¹ en la que explícitamente se denuncia el uso del factor de impacto en los procesos de evaluación de la actividad científica ha sido recibida con entusiasmo por buena parte de la comunidad investigadora. Aunque no tan evidente, la citada *Declaración*, termina cuestionando el análisis de citas en particular y la bibliometría en general para

los fines de evaluación. Aunque es cierto que se ha abusado del factor de impacto de Garfield, más allá de las propias recomendaciones de su creador y de los continuos consejos de la comunidad bibliométrica, el éxito de la *Declaración* obliga a reflexionar sobre las razones para el rechazo no ya de un indicador particular, sino de toda la estrategia basada en métricas para medir el rendimiento, visibilidad y calidad de la producción científica.

Obviamente hay que excluir del análisis al conjunto de académicos que rechazan cualquier tipo de evaluación de la actividad investigadora, puesto que no se dedican a ella. Por ejemplo, en España los incentivos económicos más importantes de los docentes universitarios están ligados a la actividad investigadora, independientemente de su carga lectiva o los medios o recursos a los que tengan acceso, lo que efectivamente dificulta o imposibilita algún tipo de producción científica. La ausencia de contrapartidas ligadas a otras actividades puede explicar el rechazo a este tipo de incentivos y su sistema de adjudicación.

“Un llamamiento para desterrar la mala bibliometría y confiar estos estudios a profesionales con experiencia”

Un segundo grupo lo constituyen aquellos científicos que han alcanzado un elevado estatus académico, incluyendo un reconocido prestigio entre sus pares locales, a partir de explotar relaciones y otros resultados intangibles, cuando no contribuciones de impacto muy localizado, pobre calidad o insignificantes. La bibliometría en estos casos puede descubrir, quizás inopinadamente para muchos de sus colegas y discípulos, las carencias y limitaciones de la carrera del investigador en cuestión. Es obvio que los interesados rechazan, incluso de manera violenta, la generalización de métodos objetivos de evaluación.

El presente ensayo pretende analizar, si además de los colectivos contrarios a las métricas, hay otras razones para explicar el rechazo de un uso más extendido de los indicadores bibliométricos. Adelantaré al lector que mi tesis es que los intereses comerciales de los productores de bases de datos bibliométricas, la desidia de la comunidad de especialistas en técnicas métricas, la mala praxis de muchos aficionados y algunos profesionales y el desconocimiento temerario de los gestores políticos y académicos pueden estar detrás de este lamentable e injustificado rechazo. De forma sintética, se intentará establecer que la causa última es la generalización en las últimas décadas de un fenómeno que doy en llamar “mala bibliometría”.

El pecado original: los índices de citas de Garfield

El monopolio durante varias décadas de las bases de datos del *ISI* (hoy *Web of science*, *WoS*, de la empresa *Thomson Reuters*) explica algunos de los problemas que enfrentamos en la actuali-

dad. Dejando a un lado los importantes sesgos de la *WoS* en la cobertura de la producción científica mundial, hay que destacar varias decisiones técnicas y comerciales clave:

- Una unidad de trabajo, la revista, que aunque útil para construir indicadores de actividad (producción), es inadecuada para crear indicadores de visibilidad o impacto. Todo ello unido a una clasificación temática de las revistas muy subjetiva, mal reflejo de las relaciones disciplinares de los títulos incluidos. Como parte de una política inmovilista que caracterizó a este sistema hasta la aparición de *Scopus*, muchas de las categorías temáticas continuaron existiendo incluso cuando era evidente su artificialidad.
- Un indicador principal (el infame factor de impacto, que crea un artefacto de citas esperadas en lugar de citas reales a los artículos, que pueden ser valores muy diferentes) común a todas las disciplinas (que en realidad tienen diferentes distribuciones y escalas) con una ventana temporal (dos años) inapropiada para muchas de las áreas y que penalizaba la estabilidad interanual de un sistema muy dinámico.
- La ausencia de un sistema de desambiguación tanto de nombre de autores como de normalización de los nombres de instituciones en las afiliaciones.
- La utilización de una matriz de citas que identificaba pares revista-revista en vez de artículo-artículo (citante-citado).
- El (posterior) sistema de cuartiles, construido a posteriori, cuando ya se había hecho, sobre los valores absolutos del indicador.

“El monopolio durante varias décadas de las bases de datos del ISI explica algunos de los problemas que enfrentamos en la actualidad”

De esos barro vinieron estos lodos

En las décadas de 1980 y 1990 el acceso a los datos de las bases del *ISI* sólo se podía realizar a través de complejas y caras estrategias en el distribuidor de bases de datos *Dialog* o mediante lentísimos lectores de cd-rom de aún más caros discos, guardados celosamente por expertos poco proclives a compartirlos y frecuentemente no actualizados. Alternativamente, una edición en papel impresa (y a veces vendida clandestinamente) de los *Journal citation reports (JCR)* se convertía en manos de inexpertos investigadores de todo el país en piedra filosofal objeto de ado-

Register for free at <https://www.scipedia.com> to download the version without the watermark



país que había obtenido de manera automática, y que mostraba una lista de científicos españoles muy citados. En muchos casos, aparte del asombro de los auditorios, los resultados causaron hilaridad entre los responsables ministeriales.

Aunque es posterior, es necesario señalar que el producto *Essential science indicators* sigue esa misma tradición de no realizar control documental sobre los resultados obtenidos de forma automática. Se pueden comprobar fácilmente los problemas de normalización de nombres (autores con 30.000 publicaciones), los fallos en las afiliaciones o las disparatadas clasificaciones y aún así sigue siendo utilizado por los amigos de la bibliometría fácil (véase por el ejemplo el ranking turco *URAP, University ranking by academic performance*).

La comunidad bibliométrica publicó durante esos años numerosos trabajos criticando abiertamente el uso abusivo del factor de impacto, con el asentimiento explícito del propio Garfield^{2, 3} presente en muchas ocasiones, pero al igual que las críticas a los problemas de cobertura, sólo como introducción a

ración. Aún hoy tan magno documento es pieza de museo, pero no se puede olvidar que algunos de ellos proponían alternativas, juiciosas unas (por ejemplo factores modificados por disciplina, o factores con ventanas más dilatadas), inviables otras, pero que nunca dieron lugar ni a bases de datos más inclusivas, ni a JCRs alternativos y ni siquiera a un acuse de recibo por parte de ISI, ya Thomson.

prolijos análisis que utilizaban tanto la denostada *ISI/Thomson* como la *ISI/Thomson* que algunos de ellos proponían alternativas, juiciosas unas (por ejemplo factores modificados por disciplina, o factores con ventanas más dilatadas), inviables otras, pero que nunca dieron lugar ni a bases de datos más inclusivas, ni a JCRs alternativos y ni siquiera a un acuse de recibo por parte de ISI, ya Thomson.

“Con la aparición de WoS, no sólo se podían filtrar los resultados por autor, afiliación, disciplina o palabra clave, sino que el análisis de citas pasó a utilizar citas reales en vez de citas esperadas”

La ausencia herramientas alternativas, complementarias o derivadas por parte tanto de la comunidad métrica como de *ISI/Thomson* es un claro símbolo del fracaso de la bibliometría, no como disciplina científica, todavía brillante, sino como

técnica incontestable y objetiva en los procesos de evaluación.

Autorías, conteos y colaboraciones

La aparición de la plataforma WoS permitió incrementar de forma considerable la cantidad y calidad de los estudios bibliométricos. No sólo se podían filtrar de forma más precisa los resultados por autor, afiliación, disciplina o palabra clave, sino que el análisis de citas pasó a utilizar citas reales en vez de citas esperadas. Aunque la desambiguación de autores seguía siendo un proceso laborioso, los trabajos micro, meso y macro eran ahora viables y al alcance de cualquiera con un poco de dedicación. Sin embargo, la supuesta facilidad generó en seguida un grave problema. Superado el modelo *Medline* de atribuir toda la autoría al primer autor/institución, y antes de que la inflación del número de firmas cuestionara el conteo fraccionado puro, la mayoría de los bibliómetros aceptaron como estándar el conteo completo, es decir asignar el 100% a todos y cada uno de los autores de los trabajos.

En el mundo de la co-autoría, perdón quiero decir de colaboración, el sucedáneo de la bibliometría fácil en un tema de cada vez mayor importancia estratégica en la definición de políticas científicas. El tratamiento descuidado de las colaboraciones asimétricas (frecuentes en Iberoamérica) provocó el rechazo de resultados poco informados, nuevamente para descrédito de la disciplina. Pero quizá la distorsión de mayor impacto a largo plazo fue que el conteo completo favoreció, aunque sea indirectamente, el aumento del número de publicaciones con múltiples firmas y también el número medio (moda) de éstas. Es necesario recordar que en muchos países de Latinoamérica el incremento porcentual de la co-autoría es muy superior al aumento de su producción bruta. Es decir, que la supuesta mejora de la actividad científica de muchas instituciones y países, de la que se ufanan políticos y gestores, no es sino un puro artefacto técnico. Cómo si no se explicaría que la producción científica española crezca de forma constante incluso después de muchos años de profunda crisis de financiación (lustros, que invalidan explicaciones basadas en condicionantes técnicos como el retraso en la publicación de resultados).



Register for free at <https://www.scipedia.com> to download the version without the watermark

El fin del monopolio *ISI* que supuso la publicación por parte de *Elsevier* de la base de datos *Scopus* parece haber traído algunas mejoras. Se cita normalmente la mayor cobertura, especialmente en ciencias sociales y humanas, aunque ya todos los expertos reconocen que esa extensión afecta sobre todo a la cola de distribución de la calidad. Aunque se incrementa el número de revistas, no aumenta de manera similar el número de trabajos, y los criterios de inclusión parecen más relajados. Pero esto tiene una consecuencia, lógica entre competidores comerciales que luchan por un mismo mercado, y que o bien ha pasado inadvertida a muchos usuarios o simplemente han preferido ignorarla: *WoS/Thomson* incrementa también su cobertura, incorporando centenares de títulos locales o regionales, lo que quiere decir que de forma artificiosa se aumenta el volumen de la producción científica a nivel mundial, pero también a nivel individual, institucional o nacional. La solución obvia de empezar a usar indicadores relativos, especialmente en los estudios temporales, no parece casar bien con ciertos intereses de expertos buscando financiación.



<http://www.timeshighereducation.co.uk/world-university-rankings>

La aparición de *Scopus* revoluciona también la batería de indicadores disponibles. Se generaliza el uso del índice *h*, sin tener en cuenta recomendaciones para matizar según edad académica o particularidades disciplinares, y *Thomson* introduce el factor de impacto de 5 años y el *Eigenfactor*, propuestas que cosechan un absoluto fracaso. Se critica su complejidad, comentario también utilizado para describir las alternativas para *Scopus* (curiosamente no desarrolladas por la propia *Elsevier*), como el *SNIP* (*source normalized impact per paper*) (realizado por el *CWTS*, *Centre for Science and Technology Studies*) o el *SJR* (de *Scimago*).

Es cuanto menos extraño que ninguno de estos indicadores claramente superiores, opinión que cualquiera con conocimientos de álgebra o simple espectador del éxito del algoritmo de *Google* suscribiría, haya sido adoptado de forma generalizada para la construcción, por ejemplo, de cuartiles de revistas. Más sorprendente aún es la reciente introducción del indicador *CPP* (citas por artículo!) por parte del *CWTS*, básicamente un remedo del factor de impacto, lo que nos hace retroceder décadas.

“La supuesta mejora de la actividad científica de muchas instituciones y países no es sino un puro artefacto técnico”

Rankings

Si hay un ejemplo paradigmático del fracaso histórico de la bibliometría éste es el éxito del llamado *Ranking de Shanghai*, un trabajo propio de un estudiante de secundaria, que sin base bibliométrica, indicadores cuestionables y nula evolución a lo largo de más de una década se ha posicionado como referente de “prestigio” en la evaluación de las universidades. Sorprende que trabajos más rigurosos (y mucho más tardíos) como el *Ranking de Leiden* o el de *SCImago* no

hayan logrado competir y que incluso compartan análisis con subproductos como el *Ranking NTU* (*National Taiwan University*) o el ya citado *URAP*.

“La aparición de *Scopus* generaliza el uso del índice *h*, sin tener en cuenta recomendaciones para matizar según edad académica o particularidades disciplinares”

Pero, a los efectos de este análisis, el ranking más interesante es posiblemente el de *Times Higher Education (THE)*. Aunque utiliza, como el ranking de *QS* y el de *US&World News*, datos de encuestas de opinión, obviamente información subjetiva y sesgada, también usa, y en un elevado porcentaje datos bibliométricos. La fuente de éstos es nada menos que *Thomson*, que no es un mero suministrador, sino que ha desarrollado específicamente indicadores para este ranking. El despropósito es inaudito y el fiasco de tal calibre que *THE* ha decidido cambiar de proveedor y pasarse a la competencia (*Scopus*): el indicador es relativo al tamaño de la producción en cada disciplina, utiliza un umbral ridículo y acepta conteos completos en áreas como la astrofísica y la física de altas energías. Ello da lugar a que, por ejemplo, universidades con un firmante entre los miles del trabajo que informa del descubrimiento del bosón de Higgs permita situarlas entre las mejores del mundo.

No obstante, los rankings han tenido la virtud de volver a poner la bibliometría en el centro del debate, incluso en países con escasa tradición en el uso de indicadores métricos. Pero no siempre las interpretaciones son correctas, lo que queda bien ilustrado con sugerencias mediocres como incrementar la producción en cuartiles inferiores, en vez de cuidar la publicación de excelencia, aumentar las co-autorías explícita e independientemente de la colaboración real, o simplemente “obligar” la normalización de las firmas y afiliaciones.

Register for free at <https://www.scipedia.com> to download the version without the watermark

Google scholar, en español Google académico

Un pecado habitual tanto dentro como fuera de la comunidad bibliométrica era identificar las carencias específicas de las fuentes (*WoS*, *Scopus*) como problemáticas intrínsecas de las disciplinas. Los análisis DAFO confundían (y aún lo hacen con la webometría y la altmetría) limitaciones coyunturales con problemas estructurales. La aparición de un tercer contendiente (*Google scholar*, *GS*) demuestra que con las herramientas adecuadas se puede evaluar objetivamente la publicación informal, las contribuciones en monografías o el creciente número de publicaciones en acceso abierto y todo ello independientemente de la disciplina, pues todas ellas (incluyendo ciencias sociales, humanidades y tecnologías, tradicionalmente maltratadas en las otras fuentes) aparecen razonablemente representadas en esta base de acceso gratuito.

Lamentablemente, aunque concediendo que *GS* tiene todavía muchas limitaciones, muchos autores están ignorando esta herramienta en sus agendas de investigación. El hecho de que algún caso se denigre o desprecie nos lleva a cuestionar si hay también intereses comerciales por parte de equipos ligados mediante acuerdos preferenciales con los otros proveedores. Incluso no sería atrevimiento cuestionar si la propia Declaración DORA no oculta asimismo intereses mercantiles.

En todo caso, señalar que la actitud ante este nuevo actor no es pro-activa, no se reclama su presencia, no se discuten sus propuestas, no se sugieren desarrollos y todo ello en un contexto, donde bajo una apreciación puramente personal, la continuidad y supervivencia de *GS* no está en absoluto garantizada.

No matéis al mensajero

Una crítica cada vez más extendida a la evaluación basada en indicadores bibliométricos es la supuesta incapacidad de éstos para identificar malas prácticas, incluyendo exceso de auto-citación, círculos de citas, falsas autorías, plagios y otras trampas varias. Ciertamente descubrir las estrategias más sofisticadas puede ser complicado, pero en mi opinión ese no es el problema principal. Como demuestran los hechos, incluso ante el flagrante “delito” el corporativismo de la academia dificulta el adecuado castigo de esas prácticas por mucho que los especialistas en métricas las identifiquen y denuncien. En el contexto de esta contribución, la buena bibliometría no sólo ha de entenderse en el sentido de publicar correctos y rigurosos estudios sino también en el de hacer públicos los datos de la

forma más transparente posible. No es frecuente la publicación de perfiles personales, de grupos o de organizaciones que permitan el escrutinio público de los datos y la posible denuncia de comportamientos poco éticos, pero son dichas herramientas y resultados los que pueden guiar a una mejor gobernanza del sistema.

“La buena bibliometría no sólo ha de entenderse en el sentido de publicar correctos y rigurosos estudios sino también en el de hacer públicos los datos de la forma más transparente posible”

A modo de conclusión

Un grupo de profesionales de la bibliometría, preocupados por el descrédito de la disciplina, motivado a nuestro parecer por la proliferación del uso poco informado y descuidado de las bases de datos bibliométricas, la poca transparencia en un entorno cada vez más comercializado y la convivencia ignorante o interesada de gestores y políticos, nos hemos venido reuniendo en los últimos años en congresos científicos en Viena, Berlín y Leiden para estudiar la situación. Se está preparando un borrador de un documento que frente al DORA recibirá el nombre de *Declaración de Leiden* y que, en espera de su redacción definitiva, hace un llamamiento para desterrar de la práctica de la bibliometría a los no profesionales con experiencia comprobada. De momento, nada que objetar.

Notas

1. American Society for Cell Biology (2013). *San Francisco Declaration on Research Assessment (DORA)*. <http://www.ascb.org/dora-old/files/SFDeclarationFINAL.pdf>
2. Garfield, Eugene (2003). “The meaning of the Impact Factor”. *Revista internacional de psicología clínica y de la salud / International journal of clinical and health psychology*, v. 3, n. 2, pp. 363-369. http://www.aepc.es/ijchp/articulos_pdf/ijchp-77.pdf
3. Garfield, Eugene (1963). “Citation indexes in sociological and historical research”. *Jasist*, v. 14, n. 4, October, pp. 289-291.

Isidro F. Aguillo

Consejo Superior de Investigaciones Científicas
Laboratorio de Cibermetría
isidro.aguillo@cchs.csic.es